

03 FILE COPY

[M][P][E][S][R][E][H]

(TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

MPsrch_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Thu Jul 2 14:43:26 1998; MasPar time 11.79 Seconds
560.820 Million cell updates/sec

Tabular output not generated.

Title: US-08-851-843-83
Description: (1-157) from-US08851843.pep
Perfect Score: 1179
Sequence: 1 FFYCTEISSVTIVYFRHDT.....FNNVLPYLFKMFVKRSCYD 157

Scoring table: PAM 150
Gap 11

Searched: 140542 seqs, 42109429 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database:
1: sp_fungi 2: sp_human 3: sp_invertebrate 4: sp_mammal
5: sp_mhc 6: sp_organelle 7: sp_phage 8: sp_plant
9: sp_bacteria 10: sp_rodent 11: sp_virus 12: sp_vertebrate
13: sp_unclassified

Statistics: Mean 43.179; Variance 84.407; scale 0.512

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Re	Score	Query Match	Length	ID	Description	Pred. No.
1	1179	100.0	884	1 Q06163	CHROMOSOME XII COSMID	1.10e-218
2	159	13.5	988	1 Q13339	TELOMERASE REVERSE TRA	2.57e-10
3	159	13.5	1031	3 Q00939	TELOMERASE SUBUNIT P12	2.57e-10
4	157	13.3	989	1 Q13338	TELOMERASE REVERSE TRA	5.50e-10
5	147	12.5	1132	2 Q14746	TELOMERASE REVERSE TRA	2.34e-08
6	147	12.5	1132	2 Q14783	TELOMERASE CATALYTIC S	2.34e-08
7	103	8.7	336	2 Q14107	PROTEIN D123.	1.25e-02
8	102	8.7	336	9 Q52228	REPLICATION-ASSOCIATED	1.25e-02
9	103	8.7	504	9 Q34044	ORF13	9.14e-02
10	102	8.7	1057	12 Q13033	RECOMBINASE	1.25e-01
11	101	8.6	349	6 Q02593	HYPOTHETICAL 40.2 KD P	1.70e-01
12	100	8.5	620	10 P97524	VERY-LONG-CHAIN ACYL-C	2.32e-01
13	99	8.4	344	9 Q28645	TRANSPPOSASE, PUTATIVE.	3.14e-01
14	99	8.4	620	10 Q35488	VERY-LONG-CHAIN ACYL-C	3.14e-01
15	98	8.3	206	9 Q27039	ATP SYNTHASE, SUBUNIT	4.26e-01
16	97	8.2	620	2 Q14975	VERY-LONG-CHAIN ACYL-C	5.76e-01
17	95	8.1	132	11 P89129	AC3 PROTEIN.	1.05e+00
18	96	8.1	132	11 P89129	TRANSPPOSASE, PUTATIVE.	7.78e-01
19	95	8.1	183	4 Q29014	ALPHA-1 ACID GLYCOPROT	1.05e+00
20	96	8.1	362	11 Q88419	SPV1-C74, COMPLETE GEN	7.78e-01

21	95	8.1	370	8 Q40093	PNIL34.	1.05e+00
22	96	8.1	530	3 Q17404	AC3.8.	7.78e-01
23	95	8.1	1280	9 Q45849	BOTULINUM NEUROTOXIN.	1.05e+00
24	94	8.0	455	1 Q08231	CHROMOSOME XV READING	1.41e+00
25	94	8.0	661	9 Q31849	YOJO PROTEIN.	1.41e+00
26	94	8.0	710	3 Q23334	ZC455.4.	1.41e+00
27	94	8.0	1547	3 Q16140	TOPOISOMERASE II (EC 5	1.41e+00
28	93	7.9	160	9 Q45198	ALTERNATIVELY SPLICED	1.89e+00
29	93	7.9	277	7 Q38606	ORF2 (FRAGMENT).	1.89e+00
30	93	7.9	517	11 Q98160	ORF2, UNKNOWN.	1.89e+00
31	93	7.9	875	9 Q31978	YOMG PROTEIN.	1.89e+00
32	93	7.9	2033	2 Q92817	ENVOLPILIN.	1.89e+00
33	92	7.8	305	9 Q32798	PUTATIVE MEMBRANE PROT	2.53e+00
34	92	7.8	305	9 Q32796	ORFA PROTEIN.	2.53e+00
35	92	7.8	516	4 Q29526	LIVER CYTOCHROME P450	2.53e+00
36	92	7.8	943	3 Q23670	K12D12.1 (FRAGMENT).	2.53e+00
37	92	7.8	1823	12 Q91062	VITELLOGENIN PRECURSOR	2.53e+00
38	91	7.7	360	3 Q26888	CATHEPSIN L-LIKE CYSTE	3.38e+00
39	91	7.7	365	3 Q23200	W06D11.4.	3.38e+00
40	91	7.7	467	8 Q04127	LETHAL LEAF-SPOT 1 (FR	3.38e+00
41	91	7.7	539	3 Q16514	T05H4.9 PROTEIN.	3.38e+00
42	90	7.6	309	9 Q06036	EPFH.	4.50e+00
43	90	7.6	532	11 Q65876	ORF2 (FRAGMENT).	4.50e+00
44	90	7.6	602	9 Q25122	GTP-BINDING MEMBRANE P	4.50e+00
45	89	7.5	1612	10 Q64399	DNA TOPOISOMERASE (ATP	5.97e+00

ALIGNMENTS

RESULT 1
ID Q06163 PRELIMINARY; PRT: 884 AA.
AC Q06163;
DT 01-NOV-1996 (TREMREL. 01, CREATED)
DT 01-NOV-1996 (TREMREL. 01, LAST SEQUENCE UPDATE)
DT 01-NOV-1996 (TREMREL. 01, LAST ANNOTATION UPDATE)
DE CHROMOSOME XII COSMID 8543.
LN 8543.12.
OS SACHAROMYCES CEREVISIAE (BAKER'S YEAST).
OC EUKARYOTA; FUNGI; ASCOMYCOTINA; HEMIASCOCYCETES.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-S288C (AB972);
RA DU Z.;
RL SUBMITTED (MAR-1995) TO EMBL/GENBANK/DBJ DATA BANKS.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-S288C (AB972);
RA JOHNSTON M., ANDREWS S., BRINKMAN R., COOPER J., DING H., DU Z.,
RA FAVELLO A., FULTON L., GATTUNG S., GRECO T., KIRSTEN J.,
RA KUCABA T., HALLSWORTH K., HAWKINS J., HILLIER L., JIER M.,
RA JOHNSON D., JOHNSTON L., LANGSTON Y., LATREILLE P., LE T.,
RA MARDIS E., MENEZES S., MILLER N., NEAN M., PAULEY A., PELUSO D.,
RA RIFKEN L., RILES L., TAICH A., TREVASKIS E., VIGNATI D.,
RA WILCOX L., WOHLDMAN P., VAUDIN M., WILSON R., WATERSTON R.;
RL SUBMITTED (MAR-1995) TO EMBL/GENBANK/DBJ DATA BANKS.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN-S288C (AB972);
RA WATERSTON R.;
RL SUBMITTED (FEB-1995) TO EMBL/GENBANK/DBJ DATA BANKS.
DR EMBL; U20618; G662136; -;
SQ SEQUENCE 884 AA; 102663 MW; 1A94320F CRC32;

Query Match 100.0%; Score 1179; DB 1; Length 884;
Best Local Similarity 100.0%; Pred. No. 1.10e-218;
Matches 157; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

DB 380 FFYCTEISSVTIVYFRHDTWNKLIITPFIVEYFKTVLVNNVCRNHNSTLSNFNSKMR 439
QY 1 FFYCTEISSVTIVYFRHDTWNKLIITPFIVEYFKTVLVNNVCRNHNSTLSNFNSKMR 60
DB 440 IIPKKNNFRIIAICPGADEEFTIYKRNHKNAIQPTQKILEYLRNKRPTSTFKIYSP 499

QY 61 IIPKSNNEFRIIAIPCRGADEEEFTIYKHNKNAIQTKILEYLRNKRPTSFYK1YSP 120

Db 500 TQIADRIKEFKORLLKFNVLPELYFMKFDVKSCYD 536
 |||||
 QY 121 TQIADRIKEFKORLLKFNVLPELYFMKFDVKSCYD 157
 |||||

RESULT 2
 ID O13339 PRELIMINARY; PRT; 988 AA.
 AC O13339;
 DT 01-JAN-1998 (TREMBLREL. 05, CREATED)
 DT 01-JAN-1998 (TREMBLREL. 05, LAST SEQUENCE UPDATE)
 DT 01-JAN-1998 (TREMBLREL. 05, LAST ANNOTATION UPDATE)
 DE TELOMERASE REVERSE TRANSCRIPTASE 1.
 GN TRT1.
 OS SCHIZOSACHAROMYCES POMBE (FISSION YEAST).
 OC EUKARYOTA; FUNGI; ASCOMYCOTINA; HEMIASCOCYCETES.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-972H-;
 RA NAKAMURA T.M., MORIN G.B., CHAPMAN K.B., WEINRICH S.L., ANDREWS W.H.,
 RA LINGNER J., HARLEY C.B., CECCH T.R.;
 RL SCIENCE 277:955-959(1997).
 DR EMBL; AF015783; G2340168;
 KW RNA-DIRECTED DNA POLYMERASE.
 SQ SEQUENCE 988 AA; 116328 MW; 39C385A7 CRC32;

Query Match 13.5%; Score 159; DB 1; Length 988;
 Best Local Similarity 27.8%; Pred. No. 2.57e-10;
 Matches 45; Conservative 36; Mismatches 68; Indels 13; Gaps 12;

Db 443 FFYITESDLNRVTYFRKDIWKLLCRPFITSMKAEKINENNVRM-DTQK-TTLPPA 500
 |||||
 QY 1 FFYCTEISSTVT-IVYFRHDTWNLITPFIVEY-FKTYL-VENNVCRNHSYTLNENHS 57
 |||||

Db 501 VIRLLPKK-NF-FRLITNLKRFLIKGSKNKKMLVSTNOTLRPVASILKHLNEESSGIP 558
 |||||
 QY 58 KRIIPKSNNEFRIIA-IPCRG-ADDEEFTIYKHNKNAIQTKILEYLRNKRPTSF 115
 |||||

Db 559 --FN-LEVYMKLLTFKKDLKH-RMFGKKYFVRIDIKSCYD 596
 |||||
 QY 116 KIYSPQIADRIKEFKORLLKFNVLPELYFMKFDVKSCYD 157
 |||||

RESULT 3
 ID O00939 PRELIMINARY; PRT; 1031 AA.
 AC O00939;
 DT 01-JUL-1997 (TREMBLREL. 04, CREATED)
 DT 01-JUL-1997 (TREMBLREL. 04, LAST SEQUENCE UPDATE)
 DT 01-JAN-1998 (TREMBLREL. 05, LAST ANNOTATION UPDATE)
 DE TELOMERASE SUBUNIT P123.
 OS EUCALYPTUS AEDICULATUS.
 OC EUKARYOTAE; MITOCHONDRIAL EUKARYOTES; ALVEOLATA; CILIOPHORA;
 OC HYPOTRICHIS; EUPLOTIDA; EUPLOTES.
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE; 97274210.
 RA LINGNER J., HUGHES T.R., SHEVCHENKO A., MANN M., LUNDBLAD V.,
 RA CECCH T.R.;
 RL SCIENCE 276:561-567(1997).
 DR EMBL; U95964; G2072336;
 SQ SEQUENCE 1031 AA; 122562 MW; 21A885CD CRC32;

Query Match 13.5%; Score 159; DB 3; Length 1031;
 Best Local Similarity 24.7%; Pred. No. 2.57e-10;
 Matches 40; Conservative 39; Mismatches 71; Indels 12; Gaps 10;

Db 455 FFVYTEQKSYKTYIRKNIWDVIMKMSIADLKETLAQVE-KEVEEKKSLGTAPGK 513
 |||||
 QY 1 FFYCTEISSTVTIYF-RHDTWNLKLPITPFIVEYFKTYLVNNVCRNHSYTLN-NFNHKS 58
 |||||

Db 514 LRLIPKKT--FRPIMTNKKIYNSDKTT-KLTNTKLNLSHMLKTLKNRMFKDPFG- 569
 |||||

QY 59 KRIIPKSNNEFR-IIAIPCRGADEEEFTIYKHNKNAIQTKILEYLRNKR-PTSFTK 116

Db 570 -FAVENYDDVKKYEE-FVCWKQVQKLPFATMDIEKCYD 609
 |||||
 QY 117 IYSPQIADRIKEFKORLLKFNVLPELYFMKFDVKSCYD 157
 |||||

RESULT 4
 ID O13338 PRELIMINARY; PRT; 989 AA.
 AC O13338;
 DT 01-JAN-1998 (TREMBLREL. 05, CREATED)
 DT 01-JAN-1998 (TREMBLREL. 05, LAST SEQUENCE UPDATE)
 DT 01-JAN-1998 (TREMBLREL. 05, LAST ANNOTATION UPDATE)
 DE TELOMERASE REVERSE TRANSCRIPTASE 1.
 GN TRT1.
 OS SCHIZOSACHAROMYCES POMBE (FISSION YEAST).
 OC EUKARYOTA; FUNGI; ASCOMYCOTINA; HEMIASCOCYCETES.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-972H-;
 RA NAKAMURA T.M., MORIN G.B., CHAPMAN K.B., WEINRICH S.L., ANDREWS W.H.,
 RA LINGNER J., HARLEY C.B., CECCH T.R.;
 RL SCIENCE 277:955-959(1997).
 DR EMBL; AF015783; G2340169;
 KW RNA-DIRECTED DNA POLYMERASE.
 SQ SEQUENCE 989 AA; 116456 MW; FDE74202 CRC32;

Query Match 13.3%; Score 157; DB 1; Length 989;
 Best Local Similarity 28.8%; Pred. No. 5.50e-10;
 Matches 47; Conservative 36; Mismatches 66; Indels 14; Gaps 13;

Db 443 FFYITESDLNRVTYFRKDIWKLLCRPFITSMKAEKINENNVRM-DTQK-TTLPPA 500
 |||||
 QY 1 FFYCTEISSTVT-IVYFRHDTWNLITPFIVEY-FKTYL-VENNVCRNHSYTLNENHS 57
 |||||

Db 501 VIRLLPKK-NF-FRLITNLKRFLIKGSKNKKMLVSTNOTLRPVASILKHLNEESSGI 558
 |||||
 QY 58 KRIIPKSNNEFRIIA-IPCRGADEEEFTIYKHNKNAIQTKILEYLRNKRPTSF 114
 |||||

Db 559 P--FN-LEVYMKLLTFKKDLKH-RMFGKKYFVRIDIKSCYD 597
 |||||
 QY 115 TKIYSPQIADRIKEFKORLLKFNVLPELYFMKFDVKSCYD 157
 |||||

RESULT 5
 ID O14746 PRELIMINARY; PRT; 1132 AA.
 AC O14746;
 DT 01-JAN-1998 (TREMBLREL. 05, CREATED)
 DT 01-JAN-1998 (TREMBLREL. 05, LAST SEQUENCE UPDATE)
 DT 01-JAN-1998 (TREMBLREL. 05, LAST ANNOTATION UPDATE)
 DE TELOMERASE REVERSE TRANSCRIPTASE.
 GN HTT.
 OS HOMO SAPIENS (HUMAN).
 OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;
 OC EUTHERIA; PRIMATES.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE-KIDNEY;
 RA NAKAMURA T.M., MORIN G.B., CHAPMAN K.B., WEINRICH S.L., ANDREWS W.H.,
 RA LINGNER J., HARLEY C.B., CECCH T.R.;
 RL SCIENCE 277:955-959(1997).
 DR EMBL; AF015950; G2330017;
 KW RNA-DIRECTED DNA POLYMERASE.
 SQ SEQUENCE 1132 AA; 126996 MW; 2DFBEDF3 CRC32;

Query Match 12.5%; Score 147; DB 2; Length 1132;
 Best Local Similarity 26.1%; Pred. No. 2.34e-08;
 Matches 43; Conservative 37; Mismatches 71; Indels 14; Gaps 14;

Db 560 FFVYTTTQKLNFFYRKSVNSKLSIGIRQHLKRVQLSELSEAEVROHRE-ARPAALT 618
 |||||
 QY 1 FFYCTEISSTVTIYF-RHDTWNLKLPITPFIVEYFKTY-LVE-NNV-CRNHNSYTLNENH 56
 |||||

73 Xa-FILE COPY

MPARCH_PP protein - protein database search, using Smith-Waterman algorithm

Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (C) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

MParch_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Thu Jul 2 14:14:03 1998; Maspar time 3.04 Seconds
72.151 Million cell updates/sec

Tabular output not generated.

Description: 38
Perfect Score: 38
Sequence: 1 DDFLXI 6

Scoring table: PAM 150
Gap 15

Searched: 120441 seqs, 36531193 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: 5:pir1 2:pir2 3:pir3 4:pir4 5:nrl3d

Statistics: Mean 20.538; Variance 24.115; scale 0.852

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description	Pred. No.
1	38	100.0	115	2	A46242 T-cell receptor alpha	1.96e+02
2	38	100.0	185	2	H64031 hypothetical protein	1.96e+02
3	38	100.0	245	2	H89471 conserved hypothetical	1.96e+02
4	38	100.0	256	2	S37861 nitrogen fixation pro	1.96e+02
5	38	100.0	266	2	S74573 hypothetical protein	1.96e+02
6	38	100.0	269	2	S03715 T-cell receptor alpha	1.96e+02
7	38	100.0	337	2	S38166 hypothetical protein	1.96e+02
8	38	100.0	351	2	S70137 phospho-N-acetylmuram	1.96e+02
9	38	100.0	364	2	S59885 H+-transporting ATP s	1.96e+02
10	38	100.0	366	2	S74847 conserved hypothetical	1.96e+02
11	38	100.0	403	2	S41708 cyclin B3 - chicken	1.96e+02
12	38	100.0	403	2	S41708 acid phosphatase (EC	1.96e+02
13	38	100.0	412	2	JN0319 genome polyprotein -	1.96e+02
14	38	100.0	415	2	S04885 hypothetical protein	1.96e+02
15	38	100.0	459	2	S76712 hypothetical protein	1.96e+02
16	38	100.0	464	2	C70122 thiophene and furan o	1.96e+02
17	38	100.0	462	2	A05026 hypothetical protein	1.96e+02
18	38	100.0	494	2	B64642 cell division protein	1.96e+02
19	38	100.0	499	2	S41599 cytochrome P450 77A1	1.96e+02
20	38	100.0	531	2	F84215 N-utilization substan	1.96e+02
21	38	100.0	884	2	S53396 hypothetical protein	1.96e+02
22	38	100.0	1118	2	S44641 F37A4.4 protein - Cae	1.96e+02
23	38	100.0	2230	1	GNNYSA genome polyprotein -	1.96e+02

24	38	100.0	2368	2	S46005	ESR1 protein - yeast	1.96e+02
25	37	97.4	105	2	JL0083	T-cell receptor delta	3.03e+02
26	37	97.4	108	2	JH0336	T-cell receptor alpha	3.03e+02
27	37	97.4	155	2	S49755	adenine phosphoribosy	3.03e+02
28	37	97.4	255	2	G69805	hypothetical protein	3.03e+02
29	37	97.4	285	2	S60598	collagen alpha chain	3.03e+02
30	37	97.4	293	1	RGBY54	translation activator	3.03e+02
31	37	97.4	331	2	S72868	hypothetical protein	3.03e+02
32	37	97.4	337	2	C64434	hypothetical protein	3.03e+02
33	37	97.4	378	2	H69505	conserved hypothetical	3.03e+02
34	37	97.4	380	1	RGHUA1	GTP-binding regulator	3.03e+02
35	37	97.4	394	1	RGBOGA	GTP-binding regulator	3.03e+02
36	37	97.4	394	1	RGHYA2	GTP-binding regulator	3.03e+02
37	37	97.4	395	1	RGHUA2	GTP-binding regulator	3.03e+02
38	37	97.4	419	2	S34421	GTP-binding regulator	3.03e+02
39	37	97.4	473	2	S49939	probable membrane pro	3.03e+02
40	37	97.4	494	2	S24387	H+-transporting ATPas	3.03e+02
41	37	97.4	536	2	A36395	spore wall maturation	3.03e+02
42	37	97.4	676	2	F69394	H+-transporting ATP s	3.03e+02
43	37	97.4	937	2	A24878	DNA-directed DNA poly	3.03e+02
44	37	97.4	1175	2	F64489	hypothetical protein	3.03e+02
45	37	97.4	3092	2	S46009	GTPase-activating pro	3.03e+02

ALIGNMENTS

RESULT ENTRY TITLE	1	A46242 T-cell receptor alpha chain variable region - mouse (fragment)	#type fragment
ORGANISM		Mus musculus	#common_name house mouse
DATE		21-Sep-1993	#sequence_revision 18-Nov-1994
ACCESSIONS		U00001	
REFERENCE		Natarajan, K.; Burstyn, D.; Zauderer, M. Proc. Natl. Acad. Sci. U.S.A. (1992) 89:8874-8878	
#authors		Major histocompatibility complex determinants select T-cell receptor alpha chain variable region dominance in a peptide-specific response.	
#journal			
#title			
#cross-references		MUID:93028369	
#accession		A46242	
#status		preliminary	
#molecule_type		nucleic acid	
#residues		1-115	#label NAT
#cross-references		GB:M84653; NID:g199287; PID:g199288	
#note		sequence extracted from NCBI backbone (NCBI:115311, NCBI:115312)	

CLASSIFICATION	#superfamily immunoglobulin V region; immunoglobulin homology
KEYWORDS	T-cell receptor
SUMMARY	#length 115 #checksum 6093

Query Match	100.0%; Score 38; DB 2: Length 115;
Best Local Similarity	83.3%; Pred. No. 1.96e+02;
Matches	5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Db	38 DDFLFI 43
Qy	1 DDFLXI 6

RESULT ENTRY TITLE	2	H64031 hypothetical protein Hii487 - Haemophilus influenzae (strain Rd KW20)	#type complete
ORGANISM		Haemophilus influenzae	#formal_name
DATE		18-Aug-1995	#sequence_revision 18-Aug-1995
ACCESSIONS		H64031	
REFERENCE		Fleischmann, R.D.; Adams, M.D.; White, O.; Clayton, R.A.; Kirkness, E.F.; Kurlavsky, A.R.; Bult, C.J.; Tomb, J.F.; Dougherty, B.A.; Merrick, J.N.; McKenney, K.; Sutton, G.;	

W P S R L H (TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

MPsrch_pp protein - protein database search, using Smith-Waterman algorithm
Run on: Thu Jul 2 14:22:02 1998; Maspar time 3.20 Seconds
Tabular output not generated. 79,803 Million cell updates/sec

Description: 1 CYDSIPR 7
Perfect Score: 59
Sequence: 1 CYDSIPR 7

Scoring table: PAM 150
Gap 15

Searched: 120441 seqs, 36531193 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: 1.pir1 2.pir2 3.pir3 4.pir4 5.nr13d

Statistics: Mean 21.467; Variance 26.929; scale 0.797

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description	Pred. No.
1	59	100.0	884	2	S53396	hypothetical protein
2	49	83.1	108	2	A48924	forkhead transcription factor
3	49	83.1	579	1	W2B34	gene 34 protein - hum
4	49	83.1	675	2	S60612	protein-tyrosine kinase
5	47	79.7	1035	2	S78199	probable maturase pro
6	46	78.0	195	2	A48190	non-actin-binding ABP
7	46	78.0	360	2	D44468	coenzyme F420 hydroge
8	46	78.0	2567	2	A49551	filamin, Mueller cell
9	45	76.3	59	5	1DXTA	venom basic protein
10	45	76.3	59	1	VTEPIA	venom basic protein
11	45	76.3	608	2	S72177	dextranase (EC 3.2.1.1)
12	45	76.3	789	1	QXB32	ox13 intron 2 protein
13	45	76.3	854	2	S17995	gene COX1 intron 1 pr
14	44	74.6	211	2	S18463	lysozyme (EC 3.2.1.17)
15	44	74.6	288	2	C56281	7alpha-cephem-methoxy
16	44	74.6	304	2	G64175	hypothetical protein
17	44	74.6	340	1	B55973	transcription factor
18	44	74.6	349	1	I51739	transcription factor
19	44	74.6	358	1	I51735	transcription factor
20	44	74.6	358	1	A55973	transcription factor
21	44	74.6	359	1	I51734	transcription factor
22	44	74.6	363	1	C55973	transcription factor
23	44	74.6	447	1	A4ECRS	argininosuccinate syn

77

FILE COPY

24 44 74.6 455 2 G01923 KIR (cl-5) NK recepto 3.43e+01
25 44 74.6 877 2 C64809 yb9B protein - Escher 3.43e+01
26 44 74.6 4639 2 A54794 dynein heavy chain, c 3.43e+01
27 43 72.9 274 2 S50190 phosphotransferase sy 5.34e+01
28 43 72.9 287 2 G69141 hypothetical protein 5.34e+01
29 43 72.9 364 2 S28612 catechol O-methyltran 5.34e+01
30 43 72.9 365 2 S18568 lignin-bispecific O-m 5.34e+01
31 43 72.9 366 2 S40146 catechol O-methyltran 5.34e+01
32 43 72.9 435 2 I39171 cyclin A/CDK2-associat 5.34e+01
33 43 72.9 490 2 S76169 hypothetical protein 5.34e+01
34 43 72.9 641 2 I50642 gammaFDP-B - chicken 5.34e+01
35 43 72.9 665 2 I50641 gammaFDP-A - chicken 5.34e+01
36 43 72.9 676 2 I50643 gammaFDP-C - chicken 5.34e+01
37 43 72.9 2368 2 S46005 ESKI protein - yeast 5.34e+01
38 43 72.9 2616 2 A57096 nudel protein precurs 5.34e+01
39 43 72.9 3791 1 YGPLV8 alpha-aminoadipyl-cys 5.34e+01
40 42 71.2 199 2 S23379 desiccation stress-in 8.26e+01
41 42 71.2 322 2 S23054 SPl protein - fruit f 8.26e+01
42 42 71.2 322 2 S23053 SPl protein - fruit f 8.26e+01
43 42 71.2 364 2 S36403 catechol O-methyltran 8.26e+01
44 42 71.2 516 2 S56604 hypothetical 58.0K pr 8.26e+01
45 42 71.2 638 1 KQRTPL plasma kallikrein (EC 8.26e+01

ALIGNMENTS

RESULT 1
ENTRY #S53396 #type complete
TITLE hypothetical protein YLR318w - yeast (Saccharomyces cerevisiae)
ALTERNATE_NAMES hypothetical protein L8543.12
ORGANISM #formal_name Saccharomyces cerevisiae
DATE 05-May-1995 #sequence_revision 01-Sep-1995 #text_change 06-Feb-1998

ACCESSIONS S53396
REFERENCE S53396
#authors Du, Z.
#submision Submitted to the EMBL Data Library, February 1995
#description The sequence of S. cerevisiae cosmid 8543.
#accession S53396
#molecule_type DNA

#residues 1-884 #label DUZ
#cross-references EMBL:U20618; NID:g2258165; PID:g662136; MIPS:YLR318w
#experimental_source strain S288C (AB972)
GENETICS
#gene SGD:EST2
#map_position 12R
#cross-references SGD:S0004310; MIPS:YLR318w
SUMMARY #length 884 #molecular-weight 102662 #checksum 7604

Query Match 100.0%; Score 59; DB 2; Length 884;
Best Local Similarity 100.0%; Pred. No. 2.09e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 534 CYDSIPR 540
Qy 1 CYDSIPR 7
|||||||

RESULT 2
ENTRY #A48924 #type fragment
TITLE forkhead transcription activator homolog (clone FXH 5-3) - human (fragment)
ORGANISM #formal_name Homo sapiens #common_name man
DATE 21-Jan-1994 #sequence_revision 18-Nov-1994 #text_change 12-Sep-1997

ACCESSIONS A48924
REFERENCE A48924
#authors Hromas, R.; Moore, J.; Johnston, T.; Socha, C.; Klemasz, M.
#journal Blood (1993) 81:2854-2859
#title Drosophila forkhead homologues are expressed in a lineage-restricted manner in human hematopoietic cells.
#cross-references MUID:93271467

Rey, M.; Reynolds, S.; Rieger, M.; Rivolta, C.; Rocha, E.; Roche, B.; Rose, M.; Sadale, Y.; Sato, T.; Scanlon, E.; Schleich, S.; Schroeter, R.; Scoffone, F.; Sekiguchi, J.; Sekowska, A.; Seror, S.J.; Serror, P.; Shin, B.S.; Soldo, B.; Sorokin, A.; Tacconi, E.; Takagi, T.; Takahashi, H.; Takemaru, K.; Takeuchi, M.; Tamakoshi, A.; Tanaka, T.; Terpstra, P.; Tognoni, A.; Tosato, V.; Uchiyama, S.; Vandenbol, M.; Vannier, F.; Vassarotti, A.; Viari, A.; Wambutt, R.; Wedler, E.; Wedler, H.; Weitzenecker, T.; Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshida, K.; Yoshikawa, H.F.; Zumstein, E.; Yoshikawa, H.; Danchin, A.

#journal Nature (1997) 390:249-256
#title The complete genome sequence of the Gram-positive bacterium Bacillus subtilis.
#accession H70052
#status preliminary; nucleic acid sequence not shown; translation not shown

GENETICS
#gene ywce
#molecule_type DNA
#residues 1-87 #label KUN
#experimental_source strain 168

KEYWORDS
#length 87 #molecular-weight 9959 #checksum 2359

Query Match 100.0%; Score 45; DB 2; Length 87;
Best Local Similarity 83.3%; Pred. No. 9.72e+01;
Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Db 42 FFYATE 47
||| ||
QY 1 FFYXTE 6

RESULT 2
ENTRY S53483 #type complete
TITLE Probable membrane protein YAR031w - yeast (Saccharomyces cerevisiae)
ORGANISM #formal_name Saccharomyces cerevisiae
DATE 05-May-1995 #sequence_revision 01-Sep-1995 #text_change 21-Nov-1997

ACCESSIONS S53483
REFERENCE S53458
#authors Bussey, H.; Keng, T.; Storms, R.K.; Vo, D.; Zhong, W.; Fortin, N.; Barton, A.B.; Kaback, D.B.; Clark, M.W.
#submission submitted to the EMBL Data Library, February 1994
#description Sequencing of chromosome I of Saccharomyces cerevisiae: analysis of the 52kbp CDC15- FLO1-PHO11-YAR074 region.

#accession S53483
#molecule_type DNA
#residues 1-298 #label BUS
#cross-references EMBL:L28920; NID:g1616966; PID:g456136; MIPS:YAR031w
GENETICS
#map_position 1R
KEYWORDS transmembrane protein.
FEATURE
114-130 #domain transmembrane #status predicted #label TM1\
141-157 #domain transmembrane #status predicted #label TM2
SUMMARY #length 298 #molecular-weight 35073 #checksum 6036

Query Match 100.0%; Score 45; DB 2; Length 298;
Best Local Similarity 83.3%; Pred. No. 9.72e+01;
Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Db 129 FFYATE 134
||| ||
QY 1 FFYXTE 6

RESULT 3
ENTRY S34811 #type complete
TITLE mating factor MAT1-2 - fungus (Cochliobolus heterostrophus)

ORGANISM #formal_name Cochliobolus heterostrophus, Bipolaris maydis
DATE 31-Dec-1993 #sequence_revision 31-Dec-1993 #text_change 08-Sep-1997
ACCESSIONS S34811; S28071
REFERENCE S34810
#authors Turgeon, B.G.; Bohlmann, H.; Cluffetti, L.M.; Christiansen, S.K.; Yang, G.; Schaefer, W.; Toder, O.C.
#journal Mol. Gen. Genet. (1993) 238:270-284
#title Cloning and analysis of the mating type genes from Cochliobolus heterostrophus.

#accession S34811
#status preliminary
#molecule_type DNA
#residues 1-343 #label TUR
#cross-references EMBL:X68398; NID:g2585; PID:g2586

GENETICS
#gene MAT1-2
#introns 162/1
CLASSIFICATION #superfamily unassigned HMG box proteins; HMG box homology
KEYWORDS DNA binding; nucleus; transcription regulation
FEATURE

SUMMARY #domain HMG box homology #label HMG1
129-203 #length 343 #molecular-weight 38319 #checksum 1452

Query Match 100.0%; Score 45; DB 2; Length 343;
Best Local Similarity 83.3%; Pred. No. 9.72e+01;
Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Db 289 FFYXTE 294
||| ||
QY 1 FFYXTE 6

RESULT 4
ENTRY I39588 #type complete
TITLE 3,4-dihydroxyphenylacetate 2,3-dioxygenase - Arthrobacter globiformis
ORGANISM #formal_name Arthrobacter globiformis
DATE 19-Jul-1996 #sequence_revision 19-Jul-1996 #text_change 19-Jul-1996

ACCESSIONS I39588
REFERENCE I39587
#authors Boldt, Y.R.; Sadowsky, M.J.; Ellis, L.B.; Que, L.
#journal J. Bacteriol. (1995) 177:1225-1232
#title A manganese-dependent dioxygenase from Arthrobacter globiformis CM-2 belongs to the major extradiol dioxygenase family.

#cross-references M01D:95173098
#accession I39588
#status preliminary; translated from GB/EMBL/DBJ
#molecule_type DNA
#residues 1-343 #label RES
#cross-references EMBL:U19817; NID:g642055; PID:g642056

GENETICS
#gene mndd
SUMMARY #length 343 #molecular-weight 38861 #checksum 7961

Query Match 100.0%; Score 45; DB 2; Length 343;
Best Local Similarity 83.3%; Pred. No. 9.72e+01;
Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Db 128 FFYETE 133
||| ||
QY 1 FFYXTE 6

RESULT 5
ENTRY E64213 #type complete
TITLE DNA topoisomerase (EC 5.99.1.2) - Mycoplasma genitalium (SGC3)
ALTERNATE_NAMES type 1 DNA topoisomerase
ORGANISM #formal_name Mycoplasma genitalium
DATE 17-Nov-1995 #sequence_revision 17-Nov-1995 #text_change

APPLICANT
COPY

FILE COPY

```
/clone="H3.4"
/clone.lib="p366, P. Hieter
/map="chromosome 12, 5.5 cm from CDC3; 10.7 cm from CDC25"
757..1977
/gene="BUD6"
/db_xref="SGD:L0002602"
757..1977
/gene="BUD6"
/codon_start=1
/db_xref="PID:6233610"
/translaton="MNGTCTPVFLRIGDKTKKCVLPPTTKPKRLLEIFERFAYSFGANS
PDYIIMPOYGVFEVLEELNLDIKEGFVIALKLENSNTIKFIDTAVKIEISNQ
NDIIRHLEMSFSAISKOTEPVLPQPLEAKHDLVQNKKDDDKTKIDQIYELGKI
KOVNINRSNINEIFNHLKVDNFSLFSKNSNRNMEKSTSHLSDTLLSKVD
DQDVIEMRKDVAERSQAPKKLETVSKDLENAQADVLKQFIDTERKPHMKKTWE
AELDKVEEOQFLTLELIDLEKALFETDLIKLCEOEKRNPSRSKNPILP
IMRPGTNOVREOVAVQSLNPDHDSRVEAIDKAEMWEMERKLKATNEFDDELENF
VGNLNKSKGGLK"
BASE COUNT 978 a 456 c 541 g 775 t
ORIGIN
```

```
Query Match 20.3%; Score 535; DB 19; Length 2750;
Best Local Similarity 98.9%; Pred. No. 0.00e+00;
Matches 546; Conservative 0; Mismatches 5; Indels 1; Gaps 1;

Db 2194 GCTTCGAAGTGTGAATTCCTTCTCAAAAGATGATATATCTTTAAATTTGATGTG 2253
Cp 2631 GCTTCGAAGTGTGAATTCCTTCTCAAAAGATGATATATCTTTAAATTTGATGTG 2572
Db 2254 TTTGAAGATAGGCTTTCACAAAATCCATTCAATATGGTGAATCGTACCTCATCTCGATT 2313
Cp 2571 TTTGAAGATAGGCTTTCACAAAATCCATTCAATATGGTGAATCGTACCTCATCTCGATT 2512
Db 2314 AAAGATCACATTTGGTAAATTTGGACACCGCTGACTGTCTTCAATGATCGTGTGAAG 2373
Cp 2511 AAAGATCACATTTGGTAAATTTGGACACCGCTGACTGTCTTCAATGATCGTGTGAAG 2452
Db 2374 AACGAATCAATTCGATATTTGACGTAACATTAATGATAGATCTTAAAGACGATTTA 2433
Cp 2451 AACGAATCAATTCGATATTTGACGTAACATTAATGATAGATCTTAAAGACGATTTA 2392
Db 2434 TAAATTCGGAATTTCTTTTACAAATGATCAATTTTCATGAGACGCTGTGTGTGA 2493
Cp 2391 TAAATTCGGAATTTCTTTTACAAATGATCAATTTTCATGAGACGCTGTGTGTGA 2332
Db 2494 TTTAAATTTGTCAATTTGTTTATAGGAGATTTCTAGTGTATAACAGCGCTATTAAACTT 2553
Cp 2331 TTTAAATTTGTCAATTTGTTTATAGGAGATTTCTAGTGTATAACAGCGCTATTAAACTT 2272
Db 2554 CGGAATATCCCTTTACTAGATTTGACGGAATGGAATATGGAATATTTCAATGCTGTAATGT 2613
Cp 2271 CGGAATATCCCTTTACTAGATTTGACGGAATGGAATATGGAATATTTCAATGCTGTAATGT 2212
Db 2614 TTCCAAATCCCAATTTCTTTTAAATAATGTCATTCGACAAATTTGAATACCGTATCA 2673
Cp 2211 TTCCAAATCCCAATTTCTTTTAAATAATGTCATTCGACAAATTTGAATACCGTATCA 2152
Db 2674 TCATCTGATTTGGAGCTTACGCTAAATAATTTGTCTCTATTGCGCTTTCCGATATATTT 2733
Cp 2151 TCATCTGATTTGGAGCTTACGCTAAATAATTTGTCTCTATTGCGCTTTCCGATATATTT 2093
Db 2734 CTGAATCCCGC 2745
Cp 2092 TTGAATCCCGC 2081
```

```
RESULT 3 166494 7218 bp DNA PAT 23-DEC-1997
LOCUS DEFINITION Sequence 14 from patent US 5670367.
ACCESSION 166494
NID 92724471
KEYWORDS Unknown.
SOURCE
```

```
Db 13406 GTGGGTTGAAGATAAGTGTACATTAGAGAAGATGCTTTTTTCAGGCTCTAGTTATC 13465
Qy 1861 GTGGGTTGAAGATAAGTGTACATTAGAGAAGATGCTTTTTTCAGGCTCTAGTTATC 1920
Db 13466 TCGTCCGATCGTGTGATTTGGTGTATGACGATCTTCTGGAGTTTATPAGGAGTTAAAGC 13525
Qy 1921 TCGTCCGATCGTGTGATTTGGTGTATGACGATCTTCTGGAGTTTATPAGGAGTTAAAGC 1980
Db 13526 CAGTCTACGAGGACACATTAATTTTAAAGCTGCTGACGATTTCTTTATATATCAAC 13585
Qy 1981 CAGTCTACGAGGACACATTAATTTTAAAGCTGCTGACGATTTCTTTATATATCAAC 2040
Db 13586 AGACCAACAGCAAGTATCAATATCAAAAGCTTGCATGGCGGATTTCAAAAATATAA 13645
Qy 2041 AGACCAACAGCAAGTATCAATATCAAAAGCTTGCATGGCGGATTTCAAAAATATAA 2100
Db 13646 TCGGAAGCAATAGACAGCAAAATTTTACCGTAGCTCCCAATCAGATGATGATCGGT 13705
Qy 2101 TCGGAAGCAATAGACAGCAAAATTTTACCGTAGCTCCCAATCAGATGATGATCGGT 2160
Db 13706 TATTCAAATTTTGTCAATGACATATTTGTTAAAGATTTGGAAGTTTGAATTCAT 13765
Qy 2161 TATTCAAATTTTGTCAATGACATATTTGTTAAAGATTTGGAAGTTTGAATTCAT 2220
Db 13766 CACAATGAATATTTCCATATCCGTTTCAAAATCTAGTAAAGGATATTTGGAAGTTTAT 13825
Qy 2221 CACAATGAATATTTCCATATCCGTTTCAAAATCTAGTAAAGGATATTTGGAAGTTTAT 2280
Db 13826 AGCGTGTTTAACAATAGATCTTATTAACAATTTGACACAAATTTAAATTCACAAA 13885
Qy 2281 AGCGTGTTTAACAATAGATCTTATTAACAATTTGACACAAATTTAAATTCACAAA 2340
Db 13886 CACCGTCTCATGCAATTTGATCATGTTGTAAGAACATTTCCGGAATGTTATTAATCTGC 13945
Qy 2341 CACCGTCTCATGCAATTTGATCATGTTGTAAGAACATTTCCGGAATGTTATTAATCTGC 2400
Db 13946 TTTAAGGATCTATCAATTAATTTACGCAAAATATGCAATTTCAATTTCTTACAACG 14005
Qy 2401 TTTAAGGATCTATCAATTAATTTACGCAAAATATGCAATTTCAATTTCTTACAACG 2460
Db 14006 CATCATTTGAATGACAGTACGAGGTTGTCCTCAATTTAGAAATGATGATCTTTAATCGAGTA 14065
Qy 2461 CATCATTTGAATGACAGTACGAGGTTGTCCTCAATTTAGAAATGATGATCTTTAATCGAGTA 2520
Db 14066 TGAGTACGATTTCAATTTGAATTTGTAAGAACATTTCCGGAATGTTATTAATCTGC 14125
Qy 2521 TGAGTACGATTTCAATTTGAATTTGTAAGAACATTTCCGGAATGTTATTAATCTGC 2580
Db 14126 ATTTAAGATTAATATCATCTTTTGAAGAACATTTCAACACTTGCACG 14176
Qy 2581 ATTTAAGATTAATATCATCTTTTGAAGAACATTTCAACACTTGCACG 2631
```

```
2 YSCPROT 2750 bp DNA PLN 26-JUL-1996
LOCUS Saccharomyces cerevisiae (clone H3.4) BUD6 gene, complete cds.
ACCESSION L38903
NID 91464752
KEYWORDS Saccharomyces cerevisiae
SOURCE Saccharomyces cerevisiae (clone library: p366, P. Hieter) DNA.
ORGANISM Saccharomyces cerevisiae
DEFINITION Eukaryote; mitochondrial eukaryotes; Fungi; Ascomycota;
Hemiascomycetes; Saccharomycetales; Saccharomycetaceae;
Saccharomycetes.
REFERENCE 1 (bases 1 to 2750)
AUTHORS Zahner, J.E. and Pringle, J.R.
TITLE Sequencing and characterization of BUD6
JOURNAL Unpublished (1995)
FEATURES location/Qualifiers
1..2750
/db_xref="taxon:4932"
/organism="Saccharomyces cerevisiae"
```


